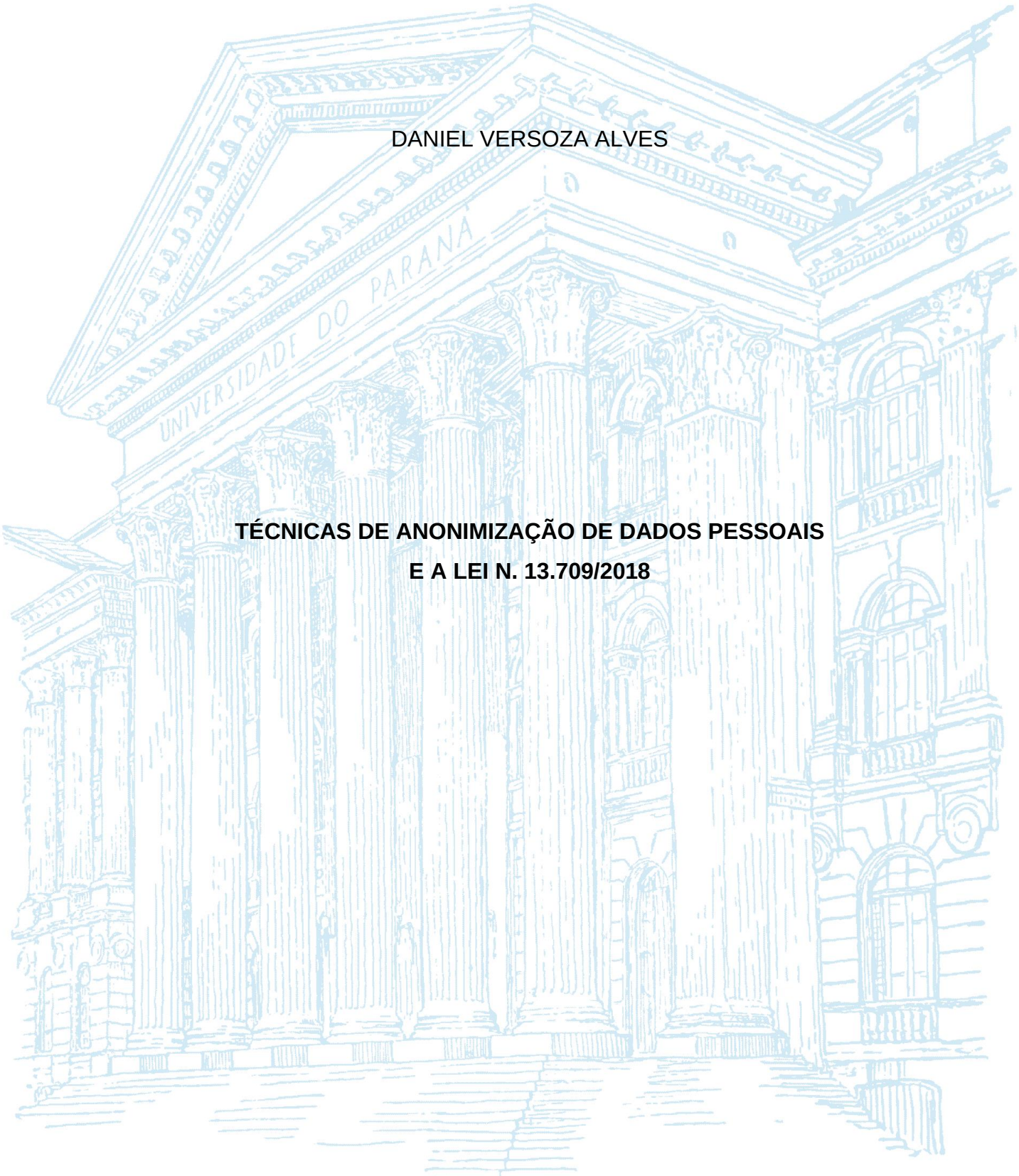
UNIVERSIDADE FEDERAL DO PARANÁ

DANIEL VERSOZA ALVES



TÉCNICAS DE ANONIMIZAÇÃO DE DADOS PESSOAIS E A LEI N. 13.709/2018

CURITIBA

2021

DANIEL VERSOZA ALVES

TÉCNICAS DE ANONIMIZAÇÃO DE DADOS PESSOAIS E A LEI N. 13.709/2018

Trabalho de Conclusão de Curso apresentado ao Curso de Direito da Faculdade de Direito da Universidade Federal do Paraná como requisito parcial à obtenção do grau de bacharel em Direito

Orientadora: Profª. Drª. Marília Pedroso Xavier

CURITIBA

2021

TERMO DE APROVAÇÃO

Técnicas de Anonimização de dados pessoais e a Lei n. 13.709/2018

DANIEL VERSOZA ALVES

Monografia aprovada como requisito parcial para obtenção de Graduação no Curso de Direito, da Faculdade de Direito, Setor de Ciências jurídicas da Universidade Federal do Paraná, pela seguinte banca examinadora:

MARILIA PEDROSO XAVIER

Assinado de forma digital por MARILIA PEDROSO XAVIER
Dados: 2021.02.26 09:33:46 -03'00'

Marília Pedroso Xavier
Orientador

Gladimir Adriani Poletto

Assinado de forma digital por Gladimir Adriani Poletto
DN: cn=Gladimir Adriani Poletto, o, ou, email=poletto@poletto.adv.br, c=BR
Dados: 2021.02.26 15:25:31 -03'00'

Gladimir Adriani Poletto
1º Membro

RODRIGO LUIS KANAYAMA

Assinado de forma digital por RODRIGO LUIS KANAYAMA
Dados: 2021.02.26 09:11:29 -03'00'

Rodrigo Luís Kanayama
2º Membro

AGRADECIMENTOS

A complexidade da vida pode ser resumida em dois pilares fundamentais: o estudo, compreendido por aquilo que aprendemos e conhecemos do mundo; e os relacionamentos, representados pelas pessoas que passam e ficam em nossas vidas.

Quanto ao primeiro pilar, estendo um especial agradecimento à Universidade Federal do Paraná, seus professores – e aqueles que se tornaram meus professores –, seu corpo técnico, auxiliares, terceirizados e demais responsáveis que puderam me garantir tantos anos de estudo público, gratuito e de qualidade.

Agradeço também à professora Marília Xavier, por ter sido minha orientadora não apensas neste trabalho, mas em tantas outras etapas fundamentais à minha formação, bem como por ter desempenhado tal função com dedicação e amizade.

Agradeço ao escritório Poletto & Possamai, que diariamente é palco de grandes experiências e ensinamentos em minha jornada profissional, bem como a todos amigos que pude lá conhecer e levarei para sempre.

Neste ponto, agradeço especialmente ao Gladimir Poletto, que além de ser um importante referencial de empreendedorismo e liderança, tenho como um grande amigo que tem sido responsável por ensinamentos que vão para além do Direito.

Quanto ao segundo pilar, a família é fundamento do que somos. Registro assim o carinho que sinto pelas famílias Versoza e Alves que muito me ensinaram e construíram quem sou, em especial suas matriarcas, Vó Ivanilde e Vó Tereza.

Sem dúvidas, dos amores que podemos receber durante a vida, aquele vindo dos pais é o mais especial. Não existe outro que consiga demonstrar ao mesmo tempo um interesse tão grande e genuíno na nossa felicidade. Por estarem sempre comigo e serem meu refúgio e porto seguro, agradeço aos meus pais, Joaquim e Renata.

Agradeço ainda ao meu irmão, Rafael, que apesar de ser bastante insuportável às vezes, é um grande companheiro de jornada e está comigo me dando apoio quando preciso.

Por fim, agradeço especialmente à Mariana Valentim, por estar sempre ao meu lado, nos momentos felizes e naqueles nem tanto, por ser minha pequena grande conselheira e voz da consciência, por ter sido minha companheira na faculdade, e hoje ser minha companheira de vida. De fato, como ela disse, não se chega a lugar nenhum sozinho, e a gratidão talvez seja o sentimento que melhor nos lembra cotidianamente dessa máxima.

If you want to keep a secret,
you must also hide it from yourself.
- George Orwell, 1984

RESUMO

O presente estudo tem como objetivo analisar o panorama de proteção de dados pessoais no que tange às técnicas de anonimização de dados no Brasil, bem como apresentar as técnicas mais utilizadas mundialmente. Primeiramente, verifica-se que o Brasil não conta com a maturidade doutrinária necessária em relação ao tema, na contramão do cenário internacional de proteção de dados, o que eleva o grau de incerteza e prejudica na implementação das técnicas. Na sequência, a anonimização de dados pessoais será esquadrinhada a partir de um ponto de vista jurídico, momento em que serão apresentadas definições conceituais necessárias à exploração do tema, bem como os *trade-offs* e riscos envolvidos nos processos de anonimização. Com isso, serão afinal apresentadas as principais técnicas de anonimização de dados pessoais, como a generalização, supressão e aleatorização, com necessários esclarecimentos quanto a da diferença entre tais processos e a pseudonimização. Finalmente, verificar-se-á que a doutrina jurídica brasileira demanda analisar o tema com maior profundidade, buscando, juntamente com os agentes de tratamento de dados, implementar as técnicas de anonimização em maior escala, visto que são o ponto de equilíbrio ideal entre a garantia de utilidade para os agentes de tratamento e a privacidade dos titulares dos dados pessoais.

Palavras-Chave: privacidade; proteção de dados; técnicas de anonimização; dados anônimos.

ABSTRACT

This study aims to analyze the Brazilian context of personal data protection in terms of anonymization techniques, exemplifying those most used worldwide. First, it is verified that Brazil is not in an adequate level of doctrinal maturity regarding the subject, contrary to the international scenario of data protection. Going on, the anonymization of personal data will be examined from a legal point of view, at which time conceptual definitions necessary for the exploration of the subject will be presented, as well as the trade-offs and risks involved in the anonymization processes. With this, the main techniques of anonymization of personal data, such as generalization, suppression and randomization will be presented, with necessary clarifications as to the difference between such processes and pseudonymization. Finally, the need for the Brazilian legal doctrine to analyze the subject in greater depth will be wide clear, which will have to seek to implement the anonymization techniques on a larger scale, since they are the ideal balance between the guarantee of usefulness for the processing agents and the privacy of the holders of personal data.

Keywords: privacy; data protection; anonymization techniques; anonymous data.

SUMÁRIO

1 INTRODUÇÃO 7
2 A ANONIMIZAÇÃO DE DADOS DESDE O PONTO DE VISTA JURÍDICO 9
2.1 CONCEITO DE DADO PESSOAL E DE DADO ANONIMIZADO 10
2.2 *TRADE-OFFS* E RISCOS DE IMPLEMENTAÇÃO ENVOLVIDOS NOS PROCESSOS DE ANONIMIZAÇÃO 13
3 TÉCNICAS DE DESIDENTIFICAÇÃO DE DADOS PESSOAIS: PSEUDONIMIZAÇÃO E ANONIMIZAÇÃO 17
3.1 NOTAS A RESPEITO DA PSEUDONIMIZAÇÃO 18
3.2 ANONIMIZAÇÃO POR SUPRESSÃO 20
3.3 ANONIMIZAÇÃO POR GENERALIZAÇÃO 21
3.3.1 k-anonimato 22
3.3.2 l-diversidade e t-proximidade 23
3.3.3 m-invariância 25
3.4 ANONIMIZAÇÃO POR ALEATORIZAÇÃO 25
3.4.1 Adição de ruído 26
3.4.2 Privacidade Diferencial 26
4 CONSIDERAÇÕES FINAIS 29

1 INTRODUÇÃO

No cenário brasileiro atual, padece-se de grande fragilidade no que se refere à categorização dos tipos de dados pessoais anonimizados e pseudonimizados. Com exceção de esparsos estudos que se preocuparam com analisar o tema com o cuidado necessário[1], poucos pesquisadores exploraram a matéria com a devida profundidade.

Possivelmente, o principal equívoco derivado de análises superficiais seja a conclusão de que técnicas de anonimização não trazem segurança jurídica enquanto não certificadas[2], constatação rasa e perigosa que desencoraja a utilização de técnicas que, ao fim e ao cabo, são fundamentais para a proteção da privacidade dos titulares de dados pessoais. Ainda, não podem ser ignoradas as equivocadas análises conjuntas de "dados sensíveis" e "dados anonimizados", que não raro são tratados como se fossem conceitos equivalentes[3].

Adicionalmente, é preciso atentar para a falta de diálogo eficiente entre agentes de tratamento de dados e pesquisadores das ciências de privacidade, desencontro comunicativo que atrasa o Brasil em relação aos países da União Europeia, por exemplo, nos quais as regulamentações de proteção da privacidade e de dados pessoais já são estudadas há vários anos[4].

Para além de afetar o meio jurídico, a falta de pesquisa sobre a matéria gera grande assimetria informacional entre os especialistas das áreas de Tecnologia da

---

[1] No direito brasileiro, vale destacar: DONEDA, Danilo; MACHADO, Diego. *Proteção de Dados Pessoais e Criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados*. In.: DONEDA, Danilo (Org.). **A regulação da criptografia no direito brasileiro**. São Paulo: Revista dos Tribunais, 2018. ISBN 978-85-203-6888-6.

[2] "*Provavelmente será necessário evoluir para o uso de soluções certificadas para evitar discussões e incertezas na aplicação de soluções de anonimização*". Ainda, na mesma obra, registra-se quanto ao apontamento de que a LGPD é inaplicável aos dados anonimizados: "*Tal apontamento da LGPD pode gerar margem a interpretação mais subjetiva e a certo grau de insegurança jurídica*". In: PINHEIRO, Patrícia Peck. **Proteção de Dados Pessoais: comentários à Lei n. 13.709/2018 (LGPD).** 2ª ed. São Paulo: SaraivaJur, 2020.ISBN 978-85-536-1748-7.

[3] A título de exemplo, cita-se trecho de obra que pretensamente aborda o assunto: "*Para ser considerado anonimizado, basta que determinado dado esteja relacionado a um titular que não possa ser identificado*". Neste ponto, verifica-se afirmação que ignora a complexidade prática da aplicação de técnicas de anonimização. Ver mais em: ARANTES, C. C. *O tratamento de dados pessoais sensíveis e os dados anonimizados*. In: **Comentários à Lei Geral de Proteção de dados e o Código de Defesa do Consumidor**. São Paulo: Editora Singular, 2019. P. 82-87.

[4] O objetivo geral das regulamentações de proteção de dados, em especial o RGPD europeu e a HIPAA estadunidense é encorajar as técnicas de mascaramento dos dados para que sejam usadas de forma geral e recorrente, como se observa neste estudo: ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 5/2014 on Anonymisation Techniques**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>. Acesso em: 19 dez. 2020.

Informação e os Juristas, conforme explica Dirceu Resende, especialista em estruturas de bancos de dados[5]:

> Infelizmente, mesmo com vários profissionais técnicos de TI ajudando a escrever essa lei, não existe nenhum termo técnico ou nada específico de TI nesse projeto [LGPD]. O texto é muito genérico e nós, profissionais de TI, ficamos à mercê de termos jurídicos e fora do nosso contexto de atuação, dificultando a interpretação do que deve ser feito para atender aos requisitos dessa lei.

Esta falta de interface entre programadores e o meio jurídico resulta em falhas de comunicação com potencial de gerar perigosos episódios de violação e vazamento de dados, como observado em notícias recentes[6].

É possível diagnosticar, inclusive, um incremento desnecessário nos custos de empresas e órgãos governamentais relativo às providências de adequação às leis de proteção de dados, visto que as companhias concentram seus esforços na adequação abstrata de contratos e redação de políticas de privacidade, mas esquecem de atentar para a proteção técnica de suas bases de dados pessoais – diligência que, comparativamente, mostra-se mais importante.

À luz de tais problemas, o objetivo do presente trabalho consiste em apresentar aos agentes de tratamentos de dados pessoais as técnicas de anonimização de dados pessoais mais utilizadas internacionalmente, a fim de torná-las mais familiares e encorajar a sua utilização em maior escala nos mais diversos bancos de dados.

Para tanto, a anonimização de dados pessoais será esquadrinhada a partir de um ponto de vista jurídico, momento em que serão apresentadas definições conceituais necessárias à exploração do tema, bem como os trade-offs e riscos envolvidos nos processos de anonimização. Ao final, serão apresentadas as principais técnicas de anonimização de dados pessoais, como a generalização, supressão e aleatorização, com necessários esclarecimentos quanto a da diferença entre tais processos e a pseudonimização.

---

[5] RESENDE, Dirceu. **Lei Geral de Proteção de Dados Pessoais (LGPDP ou LGPD) aplicada a bancos de dados SQL Server**. Publicado em: 19 mar. 2019. Disponível em: <https://www.dirceuresende.com/blog/lei-geral-de-protecao-de-dados-pessoais-lgpdp-ou-lgpd-aplicada-a-banco-de-dados-sql-server/>. Acesso em: 18 nov. 2020.

[6] PSafe dfndr lab: Vazamento em massa expõe número de CPF de milhões de brasileiros, alerta PSafe. Disponível em: <https://www.psafe.com/blog/vazamento-expoe-numero-de-cpf-de-milhoes-de-brasileiros-alerta-psafe/. Acesso em: 21 jan. 2021. Ainda, Tecnoblog: O que há no vazamento que afetou 40 milhões de CNPJs. Disponível em: < https://tecnoblog.net/404863/exclusivo-o-que-ha-no-vazamento-que-afetou-40-milhoes-de-cnpjs/>. Acesso em: 22 jan. 2021.

2 A ANONIMIZAÇÃO DE DADOS DO PONTO DE VISTA JURÍDICO

A Regulamentação Geral de Proteção de Dados (RGPD) é o principal marco regulatório da matéria, sendo chamada de "lei de ouro" sobre proteção de dados[7]. Contudo, antes mesmo da RGPD, o sistema europeu já era o mais maduro na regulação do tema, dispondo a respeito da proteção de dados pessoais na Diretiva 95/46/EC[8].

Nos Estados Unidos, a Health Insurance Portability and Accountabilty Act (HIPAA) se destaca por ser referência mundial no que diz respeito ao compartilhamento de dados, bem como os padrões de anonimização e mascaramento de dados a serem adotados pelos agentes de tratamento ao lidar com dados médios[9].

No Brasil, a Lei Geral de Proteção de Dados (LGPD)[10] é responsável por regulamentar e proteger a circulação e tratamento de dados pessoais. Diversos juristas se debruçaram sobre muitos dos seus temas[11]. Contudo, a anonimização de dados ainda é um assunto que não foi devidamente explorado. Quando analisado, a doutrina limita-se em apenas repetir a previsão legal da existência das ferramentas, mas não trata de sua implantação efetiva.

As técnicas de anonimização têm sua importância definida, de um lado, por garantir a maior privacidade dos titulares de dados pessoais no caso de trânsito e

---

[7] "*O Regulamento Europeu de Proteção de Dados Pessoais funciona como **modelo de referência** que países como o Brasil deverão levar em conta tanto na interpretação e aplicação de suas leis nacionais quanto na própria elaboração de legislação acerca da temática, em cotejo com o almejado fluxo de informações e convergências derivadas de diplomas em nível internacional.*" In: TEFFÉ, Chiara Spadaccine de; TEPEDINO, Gustavo. *Consentimento e proteção de dados pessoais na LGPD*. In: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Ed. RT, 2019. 2 ed. P. 316

[8] UNIÃO EUROPEIA (Parlamento e Conselho Europeus). **Directiva 95/46/CE**. Jornal Oficial das Comunidades Europeias, nº L 281/31. Disponível em: https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:31995L0046&from=PT. Acesso em: 20 nov. 2020.

[9] "*A HIPAA regula a proteção de informações médicas dos usuários, seja no setor hospitalar, farmacêutico ou de planos de saúde, bem como disciplinam aspectos específicos a respeito do tratamento dos dados pessoais e a necessidade de regras de privacidade.*" Ver mais em: U.S. **HIPAA: Regulation Text 45 CFR Parts 160, 162, and 164**. Disponível em: <https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/combined/hipaa-simplification-201303.pdf>. Acesso em: 23 nov. 2020.

[10] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm>. Acesso em: 17 nov. 2020.

[11] Ver mais em: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Ed. RT, 2019. 2 ed. Além disso: DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: fundamentos da lei geral de proteção de dados**. 2 ed. rev. e at. São Paulo: Ed. RT, 2019. Ainda: MALDONADO, Viviane Nóbrega; BLUM, Renato Opice (Coord.). **LGPD - Lei Geral de Proteção de Dados – Comentada**. 2 ed. São Paulo: Ed. RT, 2020.

compartilhamento de seus dados e, de outro lado, por garantir aos agentes de tratamento maior liberdade em suas operações regulares de tratamento de dados, visto que as regras de proteção de dados passam a não ser mais aplicáveis.

Em razão da lacuna regulatória e doutrinária existente no Brasil no que tange aos dados anonimizados, muito do que se irá adiante analisar resulta de estudos realizados por pesquisadores estrangeiros, o quais anotam suas observações em relação a diferentes contextos normativos.

Assim, para melhor direcionar o presente trabalho, abordaremos apenas os dois principais pontos que tangenciam o estudo da anonimização de dados: (i) os conceitos de dado pessoal e dado anonimizado; e (ii) os *trade-offs* e riscos de re-identificação inerentes às técnicas de anonimização.

2.1 CONCEITO DE DADO PESSOAL E DE DADO ANONIMIZADO

O Grupo de Trabalho de Proteção de Dados do Artigo 29º (GTPD), em seu Parecer 4/2007[12], entendeu que a expressão adotada pela Diretiva 95/46/CE, semelhante à adotada pela LGPD, tem como objetivo adotar uma noção ampla de dados pessoais, a fim de garantir em maior amplitude a proteção às pessoas naturais, tanto no caso de informações objetivas quanto subjetivas, desde que existente o vínculo relacional da informação[13].

Em resumo, dados pessoais são as informações relacionadas a uma pessoa natural identificada ou identificável[14]. Este vínculo poderá ser direto, que prontamente identificará uma pessoa (como no caso de seu nome completo, dados biométricos, e-mail ou telefone), ou indireto (como no caso de códigos identificadores, endereço, registros médicos, ou opiniões políticas).

Apesar do conceito acima ser amplamente difundido, ele não está isento de críticas. Isso porque, para além das informações pessoais direta ou indiretamente relacionadas a uma pessoa individual, qualquer informação que possa distinguir uma

---

[12] ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 4/2007 on the concept of personal data**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf >. Acesso em: 19 dez. 2020.

[13] Op cit. P. 27

[14] O artigo 5º, I, da LGPD assim define os dados pessoais: "*Para os fins desta Lei, considera-se: dado pessoal: informação relacionada a pessoa natural identificada ou identificável*".

pessoa única ou um pequeno grupo de pessoas pode ser utilizado para re-identificar dados[15], os chamados "quase-identificadores".

O grau de sensibilidade das informações, como os chamados "dados pessoais sensíveis" (artigo 5º, inciso II, LGPD), consiste em uma opinião do legislador. Esta subcategoria dos dados pessoais recebeu maior proteção jurídica em razão da observação prática dos efeitos muito mais severos decorrentes do tratamento de tais tipos de dados. Esta afirmação é corroborada por Doneda[16]:

> A própria seleção de quais seriam estes dados considerados sensíveis provém da constatação de que a circulação de determinadas espécies de informação apresentariam um elevado potencial lesivo aos seus titulares, em uma determinada configuração social.

Encontra-se nos dados sensíveis o "núcleo duro" da privacidade[17], tendo em vista que, pelo tipo e natureza de informação que trazem, apresentam dados cujo tratamento pode ensejar a discriminação ilícita ou abusiva de seu titular, devendo, por conseguinte, ser protegidos de forma mais rígida e específica. São dados especialmente sensíveis do ponto de vista dos direitos e liberdades fundamentais, cujo uso pode gerar riscos significativos para seu titular.

Por sua vez, o dado anonimizado, que não guarda qualquer relação direta com alguma categoria de dados pessoais[18], é aquele que não pode identificar uma pessoal natural após receber tratamento adequado, observados critérios objetivos de custos, tempo e técnicas razoáveis existentes à época do tratamento[19].

No que se refere à normatização dos dados anonimizados, o artigo 12 da LGPD, que delimita seu *âmbito de abrangência*, dispõe que, assim como na RGPD[20], a lei não é aplicável para o caso de tratamento de dados anonimizados:

---

[15] "*Any information that distinguishes one person from another can be used for re-identifying anonymous data.*" In: NARAYANAN, Arvind; SHMATIKOV, Vitaly. *Privacy and Security Myths and Fallacies of "Personally Identifiable Information"*. **Communications of the ACM**. Jun 2020, Vol. 53, No. 08.

[16] DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: fundamentos da lei geral de proteção de dados**. 2 ed. rev. e at. São Paulo: Thomson Reuters Brasil, 2019. P. 143.

[17] *Op cit.*

[18] Diferente do que foi explorado por alguns doutrinadores, como mencionado na NR 3.

[19] A LGPD, em seu artigo 5º, III, define o dado anonimizado como aquele: "*relativo a titular que não possa ser identificado, considerando a utilização de meios técnicos razoáveis e disponíveis na ocasião de seu tratamento"*. Ainda, em seu artigo 5º, XI, considera os processos de anonimização como: "*[a] utilização de meios técnicos razoáveis e disponíveis no momento do tratamento, por meio dos quais um dado perde a possibilidade de associação, direta ou indireta, a um indivíduo.*"

[20] A RGPD europeia, em seu considerando 26, dispôs expressamente que os dados anônimos não estão compreendidos em sua regulamentação: "*[...] Os princípios da proteção de dados não deverão, pois, aplicar-se às informações anónimas, ou seja, às informações que não digam respeito a uma*

> Art. 12. Os dados anonimizados não serão considerados dados pessoais para os fins desta Lei, salvo quando o processo de anonimização ao qual foram submetidos for revertido, utilizando exclusivamente meios próprios, ou quando, com esforços razoáveis, puder ser revertido.

Portanto, percebe-se que, diferente dos demais tipos de dados pessoais, os dados anonimizados não demandam finalidade, garantia dos princípios de proteção aos titulares dos dados, ou base legal para tratamento[21], se sustentando apenas em seus próprios atributos de segurança técnica. No contexto brasileiro ainda não foram definidos padrões técnicos adequados para anonimização, mas apenas a responsabilidade da Autoridade Nacional de Proteção de Dados (ANPD) e do Conselho Nacional de Proteção de Dados Pessoais (CNPD) quanto à disposição de tais padrões[22]:

> Art. 12. [...] § 3º A autoridade nacional poderá dispor sobre padrões e técnicas utilizados em processos de anonimização e realizar verificações acerca de sua segurança, ouvido o Conselho Nacional de Proteção de Dados Pessoais.

Em seu aspecto jurídico, a anonimização de dados tem como objetivo a irreversibilidade da identificação do titular de dados pessoais pelas informações anonimizadas, e permitir o tratamento de dados sem a incidência das regras de proteção de dados – especialmente quanto à necessidade de base legal para o tratamento ou as sanções aplicáveis no caso de vazamento. As técnicas de anonimização devem ser implementadas não apenas em grandes conjuntos de dados já coletados e existentes, mas também durante o processo de arquitetura e desenvolvimento de programas, o que é atingido por meio do que se convencionou chamar *privacy by design*. Este conceito é definido por Frazão[23] como:

---

pessoa singular identificada ou identificável nem a dados pessoais tornados de tal modo anónimos que o seu titular não seja ou já não possa ser identificado. O presente regulamento não diz, por isso, respeito ao tratamento dessas informações anónimas, inclusive para fins estatísticos ou de investigação."

[21] Os artigos 7º e 11 da LGPD trazem as bases legais para tratamento de dados pessoais ordinários e sensíveis: BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm>. Acesso em: 17 nov. 2020.

[22] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm>. Acesso em: 17 nov. 2020. Artigo 12, §3º.

[23] FRAZÃO, Ana. *Objetivos e alcance da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399. P. 118.

> [...] quando algum agente decide realizar qualquer tipo de tratamento de dados pessoais, deve pensar na privacidade em cada passo, o que inclui projeto, desenvolvimento de produtos e *softwares*, sistemas de informática, dentre outros, a fim de assegurar que a privacidade será garantida durante todo o ciclo do tratamento.

Quando adequadamente implementadas, os resultados positivos trazidos pelas técnicas de anonimização são inúmeros, visto que representam o ponto de equilíbrio ideal entre a utilidade e segurança esperada pelos agentes de tratamento e a privacidade dos titulares dos dados.

Com a anonimização de dados, por exemplo, o cenário jurídico brasileiro poderia contar com vasta jurisprudência arbitral, pela garantia da confidencialidade e sigilo das partes envolvidas, mas permitindo a divulgação das teses debatidas. Já no cenário médico, pesquisadores da saúde poderiam se valer de dados anonimizados para ter acesso a base de dados hospitalares de máximo sigilo sem que infrinjam a privacidade dos titulares de tais dados.

No entanto, a utilização adequada das técnicas de anonimização de dados depende do contexto do tratamento, da utilidade esperada[24] e da sensibilidade dos dados tratados. Por tal razão, é preciso atentar para os *trade-offs* e riscos de implementação intrínsecos a tais técnicas.

2.2 *TRADE-OFFS* E RISCOS DE IMPLEMENTAÇÃO ENVOLVIDOS NOS PROCESSOS DE ANONIMIZAÇÃO

A anonimização de dados, desde que adequadamente realizada, afasta a incidência das regras da LGPD, visto que tais dados deixam de ter qualquer relação com uma pessoa natural, deixando de ser, portanto, um dado pessoal.

Para que seja eficiente, o processo de anonimização precisa preservar os aspectos semânticos e lógicos dos dados originais, a fim de que garanta a regularidade da posterior análise exploratória. Conforme destaca Frazão[25], do ponto

---

[24] Ou seja, quais os objetivos esperados com a anonimização (tornar os dados públicos, compartilhamento com terceiros, análise estatística, traimento de máquina etc.).

[25] FRAZÃO, Ana. *Fundamentos da proteção dos dados pessoais – Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399. P. 26.

de vista econômico, os dados apenas terão algum valor caso tragam informações úteis à atividade econômica.

No entanto, há uma tensão intrínseca à economia deste "mercado de dados" entre os interesses econômicos dos controladores dos processos de tratamento de dados e os direitos de privacidade e sigilo dos seus titulares. O resultado desta difícil equação é a correlação inversamente proporcional entre o grau de privacidade adicionado à base de dados em relação à utilidade dos dados ali contidos. Na esteira da ordem econômica dos dados, Frazão[26] destaca que:

> [...] os controladores de dados criaram um sistema desenhado não para tratar os titulares de dados decentemente, mas sim para maximizar seus lucros ou colocar a inovação acima de qualquer outro valor.

Assim, os princípios de proteção de dados[27] impõem determinadas obrigações aos agentes de tratamento de dados que ao final somam conflitos em suas escolhas no que se refere aos processos de tratamento de dados, que escalam exponencialmente com o passar do tempo[28], o que Frazão optou por chamar de "*o perigoso trade-off entre os direitos dos titulares de dados e eficiências econômicas* [29].

Analisando o outro lado da moeda, no que tange aos riscos, em que pese a anonimização seja vista como um processo de total irreversibilidade de identificação, há grande dificuldade em se criar um conjunto de dados verdadeiramente anônimo[30].

---

[26] FRAZÃO, Ana. *Objetivos e alcance da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399. P. 98-99.

[27] Como a publicidade (ou transparência), a exatidão, a finalidade, o livre acesso e a segurança física e lógica. In. DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: fundamentos da lei geral de proteção de dados**. 2 ed. rev. e at. São Paulo: Thomson Reuters Brasil, 2019. ISBN 978-85-5321-957-5. Pág. 181-182.

[28] "*Graças ao desenvolvimento dos meios de armazenamento e processamento de dados, cresceria exponencialmente o custo para se manter uma informação em segredo; a privacidade ficaria mais custosa, à medida que a utilização dos dados pessoais se torna mais econômica e acessível.*" PARDOLESI, Roberto. Dalla riservatezza ala protezione dei dati personali: uma storia di evoluzione e discontinuità. Diritto ala riservatezza e circolazione dei datti personalli. Milano: Giuffrè, 2003. P. 11. *apud* DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: fundamentos da lei geral de proteção de dados**. 2 ed. rev. e at. São Paulo: Thomson Reuters Brasil, 2019. ISBN 978-85-5321-957-5. Pág. 151.

[29] FRAZÃO, Ana. *Objetivos e alcance da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399. P. 124.

[30] ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 5/2014 on Anonymisation Techniques**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>. Acesso em: 19 dez. 2020. P. 3

Exemplos recentes de algoritmos sofisticados que foram capazes de re-identificar dados de bases “anonimizadas” demonstram que o risco residual não pode ser deixado de lado[31]. Tais casos reforçam que, por mais robustas que sejam as técnicas de anonimização de dados atuais, o risco residual de re-identificação é inerente a todas elas, devendo ser visto não como óbice ao tratamento, mas um elemento que necessita de especial atenção, bem como constante reavaliação para reparar as falhas que possam aumentar este risco.

Em razão disso, a anonimização não pode ser considerada um exercício pontual, mas um trabalho recorrente que deve ser feito e reavaliado constantemente pelos responsáveis pelo tratamento de dados, especialmente com vistas à atualização tecnológica e adequação ao contexto e tipos de dados tratados[32]. De acordo com o Parecer 05/2014 do GTPD[33], são três os principais riscos inerentes aos processos de anonimização: (i) a identificação; (ii) a possibilidade de ligação; e (iii) a inferência.

O risco de *identificação* é tido como a possibilidade de isolamento de alguns ou todos os pontos de dados constantes na base de dados anonimizadas, levando a identificação de uma ou mais pessoais naturais.

Já o risco da *possibilidade de ligação*, diz respeito aos casos em que um invasor, munido de diferentes pontos de dados ou de diferentes bases de dados (anonimizadas ou não), consiga criar uma relação entre estes dados e individualize dados pessoais de uma pessoa natural, o que pode ser feito por métodos de relacionamento estatístico, análise de correlação ou outra técnica comparativa.

Por fim, o risco de *inferência* corresponde a possibilidade de se inferir o valor de determinado atributo de uma base de dados apenas a partir de um ou mais valores de outros atributos da mesma base de dados anonimizadas, podendo-se chegar à dados pessoais.

---

[31] DWORK, Cynthia; BACKSTROM, Lars; KLEINBERG, Jon. **Wherefore Art Thou R3579X? Anonymized Social Networds, Hidden Patterns, and Structural Steganography**. Ithaca: Cornell University (NY) and Microsoft Research, 2007 e NARAYANAN, Arvind; SHMATIKOV, Vitaly. **Robust De-anonymization of Large Sparse Datasets**. Texas: University of Texas at Austin, 2008.

[32] ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 5/2014 on Anonymisation Techniques**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>. Acesso em: 19 dez. 2020. P. 4.

[33] ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 5/2014 on Anonymisation Techniques**. Bruxelas: 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>. Acesso em: 19 dez. 2020. P. 12.

O Parecer GTPD reforça que uma boa técnica de anonimização é aquela que dê conta de minimizar os três riscos, mas ressalta que não há hoje uma técnica robusta o bastante que os elimine sem total perda de utilidade dos dados[34].

Em síntese, para além dos riscos de violação de privacidade, as técnicas de anonimização representam sempre uma perda de utilidade dos dados anonimizados, razão pela qual todo processo de anonimização deve ter muito claro o objetivo da utilização da técnica e a utilidade esperada dos dados anonimizados, conforme pontuado pela Personal Data Protection Commission de Singapura[35]:

> O objetivo da anonimização deve ser claro, porque a anonimização deve ser feita especificamente para o objetivo em causa. O processo de anonimização, independentemente das técnicas aplicadas, reduz a informação original no conjunto de dados em certa medida. Por isso, geralmente, à medida que a extensão da anonimização aumenta, a utilidade (ex. clareza e/ou precisão) do conjunto de dados é reduzida. Assim, a organização precisa de decidir sobre o grau de *trade-off* entre a utilidade aceitável (ou esperada) e a tentativa de reduzir o risco de re-identificação – onde o objeto dos dados é identificado a partir de dados que supostamente estariam anonimizados.

Ou seja, para um processo de anonimização eficiente, é necessário que o agente de tratamento tenha clareza do objetivo da anonimização e o grau de *trade-off* adequado entre a utilidade aceitável dos dados anonimizados e a redução do risco de violação de privacidade pela re-identificação. Os dois extremos nesse processo de avaliação são, de um lado, a não utilização de nenhuma técnica de anonimização (e manutenção integral dos dados, com grau máximo de utilidade e precisão), e, de outro lado, a anonimização completa dos dados com a supressão do conteúdo (e a consequente perda integral da utilidade de determinado atributo).

---

[34] ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 5/2014 on Anonymisation Techniques**. Bruxelas: 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>. Acesso em: 19 dez. 2020. P. 13.

[35] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 10. Item 4.a. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021

3 TÉCNICAS DE DESIDENTIFICAÇÃO DE DADOS PESSOAIS: PSEUDONIMIZAÇÃO E ANONIMIZAÇÃO

De acordo com Gavison, a privacidade é a proteção garantida de que nenhum dado analisado individualmente chame a atenção de um sujeito identificável[36]. Diversas técnicas de anonimização de dados podem ser utilizadas, mas não há previsão normativa ou regulatória a respeito da técnica mais adequada, o que será definido pela ANPD em conjunto com o CNPD[37].

Desidentificação (*de-identification*) é gênero, do qual são espécies a pseudonimização e a anonimização[38]. Enquanto a pseudonimização permite, de alguma forma, a recuperação dos dados originais por meio da reversão do processo pelo qual foi gerada, a anonimização pressupõe a quebra total desse vínculo original, pelo qual os dados originais não podem mais ser identificados ou relacionados ao seu titular[39].

Diferente da pseudonimização, a anonimização se refere ao maior e mais forte nível de desidentificação de dados, sendo definida como "[...] *uma técnica aplicada aos dados pessoais a fim de atingir uma desidentificação irreversível*"[40].

A desidentificação compreende qualquer processo pelo qual é rompido o vínculo direto estabelecido entre o dado e seu titular[41] e, por consequência disso, há maior garantia de privacidade ao titular dos dados pessoais. Apesar de ser a melhor forma de se categorizar os espectros de proteção da privacidade dos dados pessoais, optou-se pela não utilização do termo "desidentificação" no Brasil, mas apenas expressamente os termos "pseudonimização" e "anonimização".

---

[36] "*[Privacy is the] protection from being brought to the attention of others*". In: GAVISON, Ruth. Privacy and the Limits of Law. **The Yale Law Journal**, Vol. 89, No. 3 (Jan, 1980), pp. 421-471. The Yale Law Journal Company,1980. Disponível em: <http://www.jstor.org/stable/795891>. Acesso em: 15 jan. 2021.

[37] BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm>. Acesso em: 17 nov. 2020. Artigo 12, §3º.

[38] In: EMAN, Khaled El. HINTZE, Mike. **Does anonymization or de-identification require consent under the RGPD?.** Portsmouth: The International Association of Privacy Professionals, 2019. Disponível em: https://iapp.org/news/a/does-anonymization-or-de-identification-require-consent-under-the-gdpr/. Acesso em: 15 dez 2020.

[39] PRIVSEC REPORT. **Data masking: anonymization or pseudonymization?** Disponível em: https://gdpr.report/news/2017/09/28/data-masking-anonymization-pseudonymization>. Acesso em: 08 jan. 2021.

[40] ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 4/2007 on the concept of personal data**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf >. Acesso em: 19 dez. 2020. P. 7

[41] *Op Cit*.

Os processos de desidentificação, seja por anonimização ou pseudonimização, constituem formas de tratamento de dados e, portanto, para que sejam lícitos, é necessário que os dados inicialmente coletados cumpram todos os requisitos legais para sua coleta (primeiro tratamento) até que possam ser desidentificados (segundo tratamento).

Alguns conceitos são importantes para as análises que seguem. Por serem menos corriqueiros à ciência jurídica, passa-se a classificá-los[42]: (i) "*atributos*": carregam um tipo de informação de um *registro* e são sensíveis quando, agregados a *quase-identificadores*, constituem dados sensíveis (são como colunas de um banco de dados); (ii) "*registros*": cada linha de um banco de dados com seus próprios *atributos* e *valores* (são os indivíduos de um conjunto de dados); (iii) "*valor*": dentro de uma base de dados, representa o ponto de dado (como uma coordenada "x, y" relativo a um *atributo* de um *registro*); e (iv) "*quase-identificadores*"[43]: conjunto de *atributos* de um mesmo *registro* que, analisados de forma agregada, podem levar a identificação de uma pessoa natural.

Para se iniciar qualquer processo de desidentificação, é necessário identificar os dados com potencial de identificação de uma pessoa natural. Hoje, modelos complexos que mesclam inteligência artificial com técnicas de reconhecimento de entidades nomeadas (*Named Entity Recognition*[44]) podem ser utilizadas para uma varredura de bases de dados e localização de possíveis dados pessoais ou quase-identificadores.

Na sequência, devem ser escolhidas as técnicas mais adequadas para a desidentificação de cada um dos tipos de dados presente no conjunto analisado, as quais passarão a ser analisadas nos próximos tópicos.

3.1 NOTAS A RESPEITO DA PSEUDONIMIZAÇÃO

Ambos os processos de desidentificação são reconhecidos pelas leis de proteção de dados e são por elas incentivados, pois aumentam o grau de proteção da

---

[42] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 4. Item 2.2. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021.

[43] Os quase-identificadores podem ser utilizados para ligar conjuntos de dados anonimizados com outros conjuntos de dados e levar à identificação de uma pessoa.

[44] LAMPLE, Guillaume (et. al). **Neural Architectures for Named Entity Recognition**. Carnegie Mellon and Pompeu Fabra University, 2016. Disponível em: <https://arxiv.org/pdf/1603.01360.pdf>.

privacidade dos titulares. No entanto, a pseudonimização carrega graves riscos que lhe são inerentes, pois corresponde apenas a um processo de mascaramento de dados e não de anonimização[45]:

> [...] os dados sob pseudónimo não podem ser equiparados a informações anónimas, uma vez que continuam a permitir que um titular de dados seja distinguido e passível de ser ligado entre diferentes conjuntos de dados. O uso de pseudónimos é suscetível de permitir a identificação e, por conseguinte, permanece dentro do âmbito de aplicação do regime jurídico de proteção de dados.

Por consequência disso, as regulamentações de proteção de dados continuam incidindo mesmo que utilizadas técnicas de pseudonimização[46].

Um dos problemas inerentes à pseudonimização de dados, prevista no artigo 12, §2º da LGPD é a delimitação de perfis comportamentais que tenham sido construídos com base em tais dados, os quais serão considerados dados pessoais[47]:

> § 2º Poderão ser igualmente considerados como dados pessoais, para os fins desta Lei, aqueles utilizados para formação do perfil comportamental de determinada pessoa natural, se identificada.

Tais perfis são por vezes chamados pela doutrina nacional de "*representação virtual da pessoa*" ou mesmo um "*corpo eletrônico*", tendo em vista que: "*Este perfil estaria, em diversas circunstâncias, fadado a confundir-se com a própria pessoa*"[48].

Ainda que não se trate, por si só, de uma técnica de anonimização, a pseudonimização por encobrimento ou substituição de caracteres é a técnica mais

---

[45] ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 4/2007 on the concept of personal data**. Bruxelas: 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf >. Acesso em: 19 dez. 2020. P. 11

[46] Ainda que alguns pesquisadores defendam que a pseudonimização associada à exclusão dos dados originais possa ser vista como um processo de anonimização, observa-se que tais processos muito se assemelham a simples supressão ou substituição por aleatorização dos dados que venham a ser "pseudonimizados", o que são, em verdade, técnicas de anonimização. Ver mais em: DONEDA, Danilo; MACHADO, Diego. *Proteção de Dados Pessoais e Criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados*. In.: DONEDA, Danilo (Org.). **A regulação da criptografia no direito brasileiro**. São Paulo: Revista dos Tribunais, 2018. ISBN 978-85-203-6888-6.

[47] A construção de perfis comportamentais está ligada aos processos de pseudonimização que não têm o condão de eliminar ou suprimir as informações, mas manter o maior grau de qualidade possível dos dados, sendo muitas vezes suficientes para a construção de perfis, ainda que sem a presença explícita de dados pessoais.

[48] DONEDA, Danilo. **Da privacidade à proteção de dados pessoais: fundamentos da lei geral de proteção de dados**. 2 ed. rev. e at. São Paulo: Thomson Reuters Brasil, 2019. P. 152-153.

recomendada para os identificadores diretos (como nome ou e-mail), quando não suprimidos, pois garante maior utilidade para os dados[49].

De maneira semelhante, o considerando 72 do RGPD disciplina os dados formados por perfis comportamentais, compreendendo-os como dados pessoais para os fins da lei:

> (72) A definição de perfis está sujeita às regras do presente regulamento que regem o tratamento de dados pessoais, como o fundamento jurídico do tratamento ou os princípios da proteção de dados. O Comité Europeu para a Proteção de Dados criado pelo presente regulamento («Comité») deverá poder emitir orientações nesse âmbito.

Feita esta breve análise a respeito da pseudonimização, passa-se agora a expor, uma a uma, as principais técnicas de anonimização de dados pessoais.

3.2 ANONIMIZAÇÃO POR SUPRESSÃO

A anonimização pela supressão é a técnica mais radical e absoluta em termos de anonimização. No aspecto da garantia da privacidade, haverá proteção integral e irreversível de privacidade. No aspecto da utilidade, haverá perda total de utilidade dos dados. A supressão poderá ser feita com relação aos atributos de um conjunto de dados ou com relação aos seus registros.

No caso da supressão de atributos, há a remoção integral de uma seção de atributos dos dados[50]. Por ser uma técnica mais radical que prejudica absolutamente a utilidade dos dados, a supressão de atributos é apenas recomendável nos casos em que o atributo suprimido não tenha qualquer utilidade para o conjunto de dados anonimizado, ou ainda, nos casos em que nenhuma outra técnica seja capaz de anonimizá-los[51].

Em regra, deve ser aplicada a todos os atributos que identifiquem diretamente um indivíduo (como nome, CPF, e-mail, telefone, etc.), visto que têm pouca utilidade

---

[49] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 17. Item 9. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021.

[50] Se pensarmos em uma tabela, corresponderia a exclusão de determinada coluna.

[51] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 14. Item 4.a. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021

informacional e possuem grande risco de violação à privacidade no caso de vazamento ou re-identificação[52].

Uma técnica de supressão menos radical consiste na criação de um “atributo derivado” do atributo suprimido. Este atributo consiste em uma informação generalizada do dado que será suprimido e garante, de certo modo, alguma utilidade residual ao dado. Por exemplo: em um caso hipotético em que várias pessoas foram testadas quanto à presença do vírus HIV, a criação de um atributo de duração da consulta ao invés da presença de atributos de data e hora de início e de fim do exame poderá garantir a mesma utilidade ao dado, sem expor quase-identificadores de seus titulares[53].

Já no caso de supressão de registros, há a exclusão total de um ou vários registros determinados. Esta técnica é importante no caso de registros chamados “outliers”, que não satisfazem os critérios de anonimização com k-anonimato[54], conceito melhor explicado posteriormente.

É importante considerar que, assim como na supressão de atributos, a supressão de registro pode gerar graves impactos à utilidade dos dados, especialmente no caso de avaliações estatísticas como em cálculos de médias, medianas e moda, que podem ser drasticamente afetadas.

3.3 ANONIMIZAÇÃO POR GENERALIZAÇÃO

A anonimização de dados pessoais por meio da generalização consiste redução deliberada na precisão dos dados, substituindo quase-identificadores com valores menos específicos, mas semanticamente consistentes. Em regra, um processo de generalização de dados adequadamente conduzido garante grande utilidade para os dados anonimizados, em proporções semelhantes aos dados originais, mas com a garantia de anonimidade.

---

[52] SHMATIKOV, Vitaly. **k-Anonymity and Other Cluster-Based Methods**. University of Texas, CS 380S: Theory and Practice of Secure Systems. Class presentation slides. Disponível em: < https://www.cs.utexas.edu/~shmat/courses/cs380s_fall09/>. Acesso em: 23 jan. 2021.

[53] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 14. Item 4.a. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021

[54] Por exemplo: em um conjunto de dados médicos em que há dados pessoais relativos a pacientes em sua maioria entre 30 e 40 anos, caso haja um registro de um paciente de 70 anos, a desidentificação de sua idade seria muito difícil apenas pela generalização, sendo mais segura a supressão do registro.

Tradicionalmente, técnicas de generalização são utilizadas pelos censos, estatísticas públicas, entidades governamentais, entre outros, como se observa no caso de divulgação de gráficos demográficos. Isso porque, ainda que os dados coletados se refiram a pessoais naturais individuais, a divulgação estatística é feita apenas com base em faixas e valores predefinidos, com gamas de k-anonimato bastante amplas.

As precauções que devem ser tomadas no caso de generalização são, em seus extremos: a falta de agrupamento, que pode gerar identificação de alguns registros (ex.: agrupamento de endereços apenas pelo código postal que poderá levar a identificação de uma pessoa específica[55]), e o agrupamento exagerado que pode levar a perda total de utilidade dos dados (ex.: agrupar a idade de todos os indivíduos em uma faixa “entre 0 e 100 anos”, se assemelhando à supressão do atributo).

As técnicas de anonimização por generalização são mais bem utilizadas para identificadores indiretos, ou seja, dados estatísticos que possam ser agrupados, garantindo a manutenção da utilidade dos dados (ex. idade, que pode ser agrupada em níveis como “menores de 20 anos”, “entre 30 e 40 anos” e “maiores de 50 anos”)[56].

3.3.1 k-anonimato

Um dos riscos inerentes à anonimização por generalização, quando vista de forma individualizada, é deixar registros expostos à processos simples de re-identificação em razão do fato de ser o único com determinados atributos em uma base de dados[57].

A anonimização por agregação com garantia do k-anonimato se refere ao processo pelo qual diversos registros são agrupados em um único registro, em que pelo menos $k-1$ registros compartilharão os mesmos valores para todos os seus

---

[55] De acordo com estudos estatísticos conduzidos nos Estados Unidos, em razão da estrutura do código postal utilizado no país, a combinação de apenas três atributos: gênero, data de nascimento e código postal pode ser suficiente para identificar indivíduos com 87% de precisão. Mais informações disponíveis em: SWEENEY, Latanya. **Simple Demographics Often Identify People Uniquely**. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000. Disponível em: <https://dataprivacylab.org/projects/identifiability/paper1.pdf >. Acesso em: 15 dez. 2020. P. 3.

[56] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 11. Item 4.b. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021

[57] NARAYANAN, Arvind; SHMATIKOV, Vitaly. *Privacy and Security Myths and Fallacies of “Personally Identifiable Information”*. **Communications of the ACM**. Jun 2020, Vol. 53, No. 08.

atributos, garantindo a privacidade deste registro (indivíduo) em $1/k$[58]. Ou seja, a probabilidade de se re-identificar uma pessoa com base no conjunto anonimizado reduz de forma diretamente proporcional ao incremento de *k*, mas a utilidade dos dados decaí na mesma proporção[59].

Necessariamente, *k* deverá ser um número maior do que 2, para garantir que pelo menos dois indivíduos compartilhem o mesmo valor para seus atributos. Contudo, um *k* tão baixo pode ser pouco eficiente, uma vez que uma pessoa natural poderá ser identificada em todos os casos com 50% de certeza.

Um exemplo do k-anonimato é o seguinte[60]:

> Imagine um conjunto de dados específico em que k seja igual a 50 e a propriedade seja o CEP. Se observarmos qualquer pessoa desse conjunto de dados, sempre encontraremos 49 outras pessoas com o mesmo CEP. Portanto, não conseguiremos identificar nenhuma pessoa a partir do CEP dela.

O k-anonimato, no entanto, não impede ataques de inferência por homogeneidade ou ataques em que há conhecimento prévio de algum atributo sensível de terminado sujeito que conste no conjunto de dados[61], o que pode resultar na identificação desta pessoa.

3.3.2 l-diversidade e t-proximidade

Apesar de ser uma técnica muito eficiente, o k-anonimato não está livre de limitações e riscos de re-identificação[62]. Nesse sentido, as técnicas de generalização

---

[58] NARAYANAN, Arvind; SHMATIKOV, Vitaly. *Privacy and Security Myths and Fallacies of "Personally Identifiable Information"*. **Communications of the ACM**. Jun 2020, Vol. 53, No. 08.

[59] A estruturação lógica da definição consiste em: "*Considerando TP como a tabela de publicação, $QI_{TP} = (A_i, \dots, A_j)$ como o conjunto de quase-identificadores associados com a TP, $A_i, \dots, A_j \subseteq A_i, \dots, A_n$, e TP satisfaça o k-anonimato. Então, cada sequência de valores em $TP[A_x]$ aparecem com pelo menos k ocorrências em $TP[QI_{TP}]$ para $x = i, \dots j$.*" (tradução livre). In: SHMATIKOV, Vitaly. **k-Anonymity and Other Cluster-Based Methods**. University of Texas, CS 380S: Theory and Practice of Secure Systems. Class presentation slides. Disponível em: < https://www.cs.utexas.edu/~shmat/courses/cs380s_fall09/>. Acesso em: 23 jan. 2021.

[60] GOOGLE. Como o Google Anonimiza os Dados. In: **Privacidade & Termos do Google**. Disponível em: <https://policies.google.com/technologies/anonymization?hl=pt-BR>. Acesso em: 12 dez. 2020.

[61] PATEL, Twinkle; AMIN, Kiran. *A Study on k-anonymity, l-diversity, and t-closeness: Techniques of Privacy Preservation Data Publishing*. **IJIRST –International Journal for Innovative Research in Science & Technology**, Volume 6, Issue 6, Nov 2019. ISSN (online): 2349-6010. P. 4.

[62] Como ocorre nos casos em que há grande proximidade nos registros de pesquisa ou quando há necessidade de republicação dos dados. Ver mais em: XIAO, Xiaokui; TAO, Yufei. **m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets**. Hong Kong: Chinese University of Hong Kong, Department of Computer Science and Engineering, 2007.

por l-diversidade são importantes para os casos em que a recorrência de registros, ainda que anonimizados com garantia do k-anonimato, permitam a inferência de determinada situação.

Para a aplicação correta da l-diversidade, é necessário que cada atributo de valores sensíveis ou quase-identificadores estejam “bem representados” no conjunto de dados. A noção de uma boa representação pode ser interpretada de várias formas, mas a mais comum é a de que cada grupo de quase-identificadores, pelo menos $1/l$ dados deverão conter informações distintas[63]. Assim, quanto maior o valor de l, associado ao maior valor de k, a privacidade estará mais garantida.

Um dos melhores exemplos da l-diversidade é o seguinte[64]:

> Imagine que um grupo de pessoas tenha pesquisado o mesmo tópico de saúde (por exemplo, sintomas da gripe), todas ao mesmo tempo. Se analisarmos esse conjunto de dados, não conseguiremos dizer quem pesquisou o tópico, graças ao k-anonimato. No entanto, ainda poderá haver alguma preocupação em relação à privacidade, uma vez que todos compartilham do mesmo atributo de confidencialidade (ou seja, o tópico da pesquisa). Com a l-diversidade, o conjunto de dados anonimizados não incluiria apenas pesquisas sobre a gripe, mas poderia incluir também outras pesquisas para proteger ainda mais a privacidade do usuário.

Um dos riscos da l-diversidade é que, por si só, não avalia nem garante distribuição semelhante de dados sensíveis, tampouco considera a semântica de tais dados, não impede ataques de inferência ou possibilidade de ligação [65].

Classificada como um “refinamento” da l-diversidade, a t-proximidade consiste na criação de classes equivalentes de registros que garantam a distribuição de valores de forma próxima à distribuição da base de dados original[66]. Esta técnica é utilizada para aumentar a qualidade da base de dados sem que haja perda de privacidade.

No entanto, não existe por si só. Para que seja utilizada, a t-proximidade deve ser inserida no contexto de um conjunto de dados l-diverso e k-anônimo.

---

[63] MACHANAVAJJHALA, Ashwin; GEHRKE, Johannes; KIFER, Daniel; VENKITASUBRAMANIAM, Muthuramakrishnan. **l-diversity: Privacy beyond k-anonymity**. ACM Transactions on Knowledge Discovery from Data. Vol. 1, Mar, 2007. Disponível em: https://dl.acm.org/doi/abs/10.1145/1217299.1217302. Acesso em: 16 nov. 2020.

[64] GOOGLE. Como o Google Anonimiza os Dados. In: **Privacidade & Termos do Google**. Disponível em: <https://policies.google.com/technologies/anonymization?hl=pt-BR>. Acesso em: 12 dez. 2020.

[65] PATEL, Twinkle; AMIN, Kiran. *A Study on k-anonymity, l-diversity, and t-closeness: Techniques of Privacy Preservation Data Publishing*. **IJIRST –International Journal for Innovative Research in Science & Technology**, Volume 6, Issue 6, Nov 2019. ISSN (online): 2349-6010. P. 4.

[66] “t” está para “*threshold*” ou, em português, “*limite*”.

3.3.3 m-invariância

Um dos problemas das técnicas tradicionais de generalização que visam apenas ao k-anonimato, à l-diversidade e à t-proximidade, é a impossibilidade de republicação dos dados. Ou seja, após a generalização de dados estáticos e sua publicação, caso haja a necessidade de uma nova publicação da mesma base de dados considerando a inclusão, alteração ou, principalmente, a exclusão de registros pré-existentes, torna-se possível a re-identificação pela comparação entre a antiga e a nova base de dados, ou seja, há o risco da possibilidade de ligação [67].

A m-invariância consiste na inserção de grupos de dados falsos ("tuplas" falsificadas[68]) importados de tabelas anteriormente publicadas. Para garantir a segurança e invariância do processo, deve ser adicionado um atributo de "vida útil" aos registros dos diferentes conjuntos de dados publicados que resulta da junção entre os conjuntos de dados históricos e atuais.

Portanto, a aplicação do conceito de m-invariância possibilita a republicação de bases de dados anônimos sem permitir que um invasor possa inferir dados pelo relacionamento entre os dados previamente publicados[69].

3.4 ANONIMIZAÇÃO POR ALEATORIZAÇÃO

Os processos de anonimização por aleatorização constituem um grupo com diferentes espécies de técnicas que têm como finalidade a alteração da veracidade de alguns dados com o objetivo de romper a ligação entre tais dados e uma pessoa natural. Em geral, todos os atributos e valores originais são mantidos, mas com registros alterados de forma aleatória – por isso essas técnicas são por vezes chamadas de técnicas de "embaralhamento"[70].

---

[67] XIAO, Xiaokui; TAO, Yufei. **m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets**. Hong Kong: Chinese University of Hong Kong, Department of Computer Science and Engineering, 2007. P. 2.

[68] Tradução livre da expressão "*couterfeit tuples*". Em programação, *tuples* (tuplas ou énuplos) são sequências ordenadas de elementos, que podem ser definidas pela recursão de pares ordenados. Caso a sequência consista em apenas dois elementos, será chamada "dupla". Ver mais em: <https://www.thefreedictionary.com/tuple>. Acesso em: 23 jan. 2021.

[69] "The rationale of m-invariance is that, if a tuple t (from the microdata) is published several times, all its generalized hosting groups must contain the same sensitive values". In: XIAO, Xiaokui; TAO, Yufei. **m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets**. Hong Kong: Chinese University of Hong Kong, Department of Computer Science and Engineering, 2007. P. 7.

[70] Em inglês, o registro mais recorrente é de "*data shuffling*".

As técnicas de aleatorização são bastante recomendáveis nos casos de pesquisas estatísticas, quando a apuração individual dos atributos sem necessária correlação direta com seus registros seja suficiente para a análise[71].

Um dos problemas decorrentes da má utilização da aleatorização é a possibilidade de uma fácil inferência dos valores reais dos dados em pequenos conjuntos.

3.4.1 Adição de ruído

A aleatorização por adição de ruído consiste na ligeira alteração randômica “para cima” ou “para baixo” dos dados originais. Em grandes conjuntos de dados, esta alteração pouco afeta o resultado estatístico final, pois as probabilidades de variações serão pequenas, mantendo-se os dados dentro de um intervalo de variância normal, com desvio padrão reduzido.

Quanto mais ruído é adicionado, mais protegidos estão os dados, mas menos precisos, relevantes e úteis eles se tornam. Um exemplo da adição de ruído clássica é a alteração aleatória do atributo “data de nascimento” com a adição ou subtração de um valor entre 0 e 2 anos para todos os indivíduos.

Esta técnica é bastante útil no caso de quase-identificadores que, analisados em conjunto, possam resultar na identificação de uma pessoa natural, bem como no caso de grandes conjuntos de dados, pela garantia de utilidade estatística dos dados.

Um dos pontos positivos desta técnica é a possibilidade de se mensurar adequadamente o grau de privacidade em detrimento da utilidade dos dados pessoais com base na escala e força de interferência do desvio escolhido.

3.4.2 Privacidade Diferencial

A privacidade diferencial é um método de anonimização que combina a técnica da aleatorização com adição de ruído e a entrega de resultado de consultas individuais aleatoriamente diferentes.

---

[71] SINGAPORE (Personal Data Protection Commission). **Guide to Basic Data Anonymization Techniques**. Publicado em 25 jan. 2018. P. 23. Item 11. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021.

> A privacidade diferencial é responsável por endereçar o paradoxo de não entender nada sobre indivíduo algum, mas ainda assim apender informações úteis sobre uma população[72].

O cálculo comparativo entre a "utilidade" e a "privacidade" dos dados que passam pela privacidade diferencial é medida por ε (letra grega épsilon), sendo ε diretamente proporcional à precisão dos dados e, via da consequência, inversamente proporcional à privacidade.

A proposta do método da privacidade diferencial diverge dos demais casos citados anteriormente, pois não é adequada para a divulgação de dados anônimos, mas apenas para estruturas em que haja diferentes consultas a um conjunto de dados anonimizado. Apesar de divergir das demais técnicas, a privacidade diferencial também poderá complementá-las, na medida em que, a depender do contexto do tratamento de dados, o agente poderá utilizar as demais técnicas de anonimização aliada a uma entrega aleatoriamente diferenciada para cada consulta ao seu conjunto de dados.

Quando o ruído é adicionado antes da coleta dos dados é chamada de "privacidade diferencial local", ou seja, a informação coletada do usuário pode ser ou não considerada. Já quando o ruído é adicionado após a coleta dos dados, é chamada de "privacidade diferencial central".

Na esteira da privacidade diferencial, o "aprendizado federado" (ou *federated learning*), consiste na utilização de técnicas de aprendizado de máquina descentralizadas, sem a necessidade de um único banco de dados no qual há adição de ruído, mas em vários bancos difusos com aplicação de privacidade diferencial local.

Hoje, técnicas mais avançadas de privacidade diferencial aliadas ao aprendizado federado são utilizadas por grandes empresas (como Apple[73] e Google[74]) para treinamento de suas assistentes virtuais. Em ambos os casos, os dados de voz

---

[72] PILLOW, Timothy. *A Review of Synthetic Tabular Data Tools and Models**:** Anonymization methods that are revolutionizing how we share data.* **Medium**: 02 jul. 2020. Disponível em: <https://towardsdatascience.com/a-review-of-synthetic-tabular-data-tools-and-models-d83b232aae25>. Acesso em: 15 nov. 2020.

[73] BHOWMICK, Abhishek; DUCHI, John; FREUDIGER, Julien; KAPOOR, Gaurav; ROGERS, Ryan. **Protection Against Reconstruction and Its Applications in Private Federated Learning**. Apple, in collaboration with Stanford University: 2019. Disponível em: https://arxiv.org/pdf/1812.00984.pdf. Acesso em: 15 jan. 2021.

[74] MCMAHAN, Brendan; RAMAGE, Daniel. **Federated Learning: Collaborative Machine Learning without Centralized Training Data**. Google: 2017. Disponível em: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html. Acesso em: 15 jan. 2021.

e comportamento dos usuários são coletados, tratados com adição de ruído em seus próprios dispositivos e, só então, são enviados às centrais de treinamento de máquina.

Assim como existem *trade-offs* na escolha de determinados métodos de anonimização no que diz respeito à perda de utilidade dos dados anonimizados, há também custos de complexidade para criação de um sistema de privacidade, neste caso, de privacidade diferencial[75]. Por tal razão, pode fazer pouco sentido para a maior parte das empresas, ainda que seja defendida por diversos pesquisadores[76] como a técnica ideal para o futuro.

No entanto, a privacidade diferencial carrega consigo a impossibilidade de oferecer uma metodologia eficiente e de baixo custo para o compartilhamento de dados, especialmente de pequenas bases de dados, o que reforça a necessidade da construção personalizada da técnica mais adequada para a anonimização de cada categoria de dados em cada base de dados[77].

---

[75] "*Differential privacy can be complicated and timely to setup. Just as there is a tradeoff between level of anonymization and data utility, there is a tradeoff between effort and complexity of a privacy system*." In: PILLOW, Timothy. *A Review of Synthetic Tabular Data Tools and Models**:** Anonymization methods that are revolutionizing how we share data.* **Medium**: 02 jul. 2020. Disponível em: <https://towardsdatascience.com/a-review-of-synthetic-tabular-data-tools-and-models-d83b232aae25>. Acesso em: 15 nov. 2020.

[76] Como Andrew Trask e demais pesquisadores da Oxford University e da comunidade OpenMined de treinamento para o uso de inteligência artificial em aprendizagem federada e privacidade diferencial.

[77] "*Differential privacy, however, does not offer a universal methodology for data release or collaborative, privacy-preserving computation. This limitation is inevitable: privacy protection has to be built and reasoned about on a case-by-case basis.*" In: NARAYANAN, Arvind; SHMATIKOV, Vitaly. *Privacy and Security Myths and Fallacies of "Personally Identifiable Information"*. **Communications of the ACM**. Jun 2020, Vol. 53, No. 08.

4 CONSIDERAÇÕES FINAIS

Padece-se no Brasil de uma grave lacuna doutrinária entre os operadores do Direito no que diz respeito às técnicas de anonimização de dados pessoais. As principais consequências disso são o desencorajamento da utilização de tais técnicas e a assimetria informacional entre juristas e técnicos em segurança da informação.

Por decorrência disso, há ainda um receio latente dos agentes de tratamento quanto à postura que devem adotar frente aos dados pessoais que irão tratar, especialmente no tocante aqueles classificados como sensíveis.

Face a este problema, o presente trabalho buscou construir uma análise a respeito dos principais conceitos trazidos pela Lei Geral de Proteção de Dados, comparada às demais normativas de proteção de dados internacionais (especialmente a HIPAA estadunidense e a RGPD europeia), para demonstrar a importância da utilização de técnicas de anonimização como forma de garantir maior segurança e privacidade aos titulares de dados pessoais.

Além disso, foram analisados os *trade-offs* e riscos inerentes aos processos de anonimização que, invariavelmente, representam menor utilidade aos dados tratados, mas maior privacidade aos titulares dos dados, de acordo com o maior critério de segurança aplicado.

A fim de tornar mais eficiente o equilíbrio entre a utilidade dos dados e privacidade dos titulares, cumpre aos agentes de tratamento analisarem suas bases de dados, os tipos de dados tratados, o contexto em que estão inseridos e a utilidade esperada para, então, aplicar as técnicas de anonimização e pseudonimização mais adequadas.

Ainda que a pseudonimização não possa ser vista como técnica de anonimização, especialmente por permitir a simples reversão do processo e a construção de perfis das pessoas que contiverem registros na base de dados, esta deverá ser aplicada sempre que possível, principalmente no tocante aos identificadores diretos e para os casos de publicação ou compartilhamento de dados.

Por conseguinte, as técnicas de anonimização de dados podem ser desde as mais radicais e absolutas, como a supressão integral de registros ou atributos do conjunto de dados, passando pelas técnicas de generalização que, idealmente, deverão garantir o k-anonimato, a l-diversidade, a t-proximidade e, para o caso de

republicação, a m-invariância de seus dados, até as técnicas de aleatorização, com adição de ruído ou a que tem se tornado mais popular, a privacidade diferencial.

Ferramentas como Google Cloud Data Loss Prevention[78], Azure Microsoft Information Protection[79], Broadcom Symantec Data Loss Prevention[80] e Amazon Web Services Macie[81] são importantes para facilitar o acesso a técnicas de proteção à privacidade de dados de qualidade com baixo custo para diversos agentes de tratamento de dados. No entanto, por serem mecanismos padronizados e previamente desenvolvidos, o controle sobre os mecanismos de anonimização dos dados é perdido e o agente de tratamento será dependente da completude de tais ferramentas.

Por tal razão, a criação manual e personalizada de funções/algoritmos de anonimização de dados é necessária, que deve ser adequada ao contexto, à utilidade esperada, aos tipos de dados e a sua sensitividade, à base legal de tratamento dos dados, e às responsabilidades dos agentes de tratamento.

Por fim, para além da utilização das técnicas de desidentificação de dados pessoais mais sofisticadas, é essencial que um sistema de privacidade eficiente conte com procedimentos de segurança, mecanismos tecnológicos e físicos de controle e restrição de acesso às suas bases de dados, criação de protocolos de comunicação entre máquinas (antivírus, SSH, etc.), na rede (firewall, HTTP, FTP ou certificados) e interface humana (senhas, biometria, etc.), sob pena de que as técnicas de segurança lógica aplicadas restem prejudicadas por falhas físicas e humanas na preservação da base de dados.

[78] Mais informações em: https://cloud.google.com/dlp
[79] Mais informações em: https://azure.microsoft.com/en-us/services/information-protection/
[80] Mais informações em: https://www.broadcom.com/products/cyber-security/information-protection/data-loss-prevention
[81] Mais informações em: https://aws.amazon.com/pt/macie/

REFERÊNCIAS BIBLIOGRÁFICAS

ARANTES, C. C. *O tratamento de dados pessoais sensíveis e os dados anonimizados*. **Comentários à Lei Geral de Proteção de dados e o Código de Defesa do Consumidor**. São Paulo: Editora Singular, 2019.

ARTICLE 29 DATA PROTECTION WORKING PARTY. **Opinion 4/2007 on the concept of personal data**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2007/wp136_en.pdf >. Acesso em: 19 dez. 2020.

_____. **Opinion 5/2014 on Anonymisation Techniques**. Bruxelas: [s. n.], 2007. Disponível em: <https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf>. Acesso em: 19 dez. 2020.

BIONI, Bruno Ricardo. **Proteção de dados pessoais: a função e os limites do consentimento**. Rio de Janeiro: Forense, 2019. XXVII, 314 p., il, 24 cm. Inclui notas explicativas, bibliográficas, de jurisprudência e bibliografia. ISBN 9788530981686.

BHOWMICK, Abhishek; DUCHI, John; FREUDIGER, Julien; KAPOOR, Gaurav; ROGERS, Ryan. **Protection Against Reconstruction and Its Applications in Private Federated Learning**. Apple, in collaboration with Stanford University: 2019. Disponível em: https://arxiv.org/pdf/1812.00984.pdf. Acesso em: 15 jan. 2021.

BRASIL. **Lei nº 13.709, de 14 de agosto de 2018**. Lei Geral de Proteção de Dados Pessoais (LGPD). Disponível em: <http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm>. Acesso em: 17 nov. 2020.

DONEDA, Danilo (Org.). **A regulação da criptografia no direito brasileiro**. São Paulo: Revista dos Tribunais, 2018. ISBN 978-85-203-6888-6.

_____. **Da privacidade à proteção de dados pessoais: fundamentos da lei geral de proteção de dados**. 2 ed. rev. e at. São Paulo: Editora Revista dos Tribunais, 2019.

_____.; MACHADO, Diego. *Proteção de Dados Pessoais e Criptografia: tecnologias criptográficas entre anonimização e pseudonimização de dados*. **A regulação da criptografia no direito brasileiro**. São Paulo: Revista dos Tribunais, 2018. ISBN 978-85-203-6888-6.

CASCAES, Amanda Celli; ALMEIDA, Fabíola Meira de; TUTIKIAN, Priscila David Sansone (coord.). **Comentários à Lei Geral de Proteção de Dados à Luz do código de defesa do consumido**r. 1. ed. São Paulo: Singular, 2019. 400 p., 24 cm. Inclui referências. ISBN 9788553066261.

DWORK, Cynthia; BACKSTROM, Lars; KLEINBERG, Jon. **Wherefore Art Thou R3579X? Anonymized Social Networds, Hidden Patterns, and Structural Steganography**. Ithaca: Cornell University (NY) and Microsoft Research, 2007

EMAN, Khaled El. HINTZE, Mike. **Does anonymization or de-identification require consent under the RGPD?** Portsmouth: The International Association of Privacy Professionals, 2019. Disponível em: https://iapp.org/news/a/does-anonymization-or-de-identification-require-consent-under-the-gdpr/. Acesso em: 15 dez 2020.

FRAZÃO, Ana. *Fundamentos da proteção dos dados pessoais – Noções introdutórias para a compreensão da importância da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399.

_____. *Objetivos e alcance da Lei Geral de Proteção de Dados*. In.: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399.

GAVISON, Ruth. **Privacy and the Limits of Law**. The Yale Law Journal, Vol. 89, No. 3 (Jan, 1980), pp. 421-471. The Yale Law Journal Company,1980. Disponível em: <http://www.jstor.org/stable/795891>. Acesso em: 15 jan. 2021.

GRAHAM, Christopher. **Anonymisation: managing data protection risk code of practice**. Information Commissioner's Office: Wilmslow, 2012. Disponível em: https://ico.org.uk/media/1061/anonymisation-code.pdf. Acesso em: 10 jan. 2021.

GOOGLE. **Como o Google Anonimiza os Dados**. In: Privacidade & Termos do Google. Disponível em: <https://policies.google.com/technologies/anonymization?hl=pt-BR>. Acesso em: 12 dez. 2020.

LAMPLE, Guillaume (et. al). **Neural Architectures for Named Entity Recognition**. Carnegie Mellon and Pompeu Fabra University, 2016. Disponível em: <https://arxiv.org/pdf/1603.01360.pdf>.

MACHANAVAJJHALA, Ashwin; GEHRKE, Johannes; KIFER, Daniel; VENKITASUBRAMANIAM, Muthuramakrishnan. **l-diversity: Privacy beyond k-anonymity**. ACM Transactions on Knowledge Discovery from Data. Vol. 1, Mar, 2007. Disponível em: https://dl.acm.org/doi/abs/10.1145/1217299.1217302. Acesso em: 16 nov. 2020.

MALDONADO, Viviane Nóbrega; BLUM, Renato Opice (Coord.). **LGPD - Lei Geral de Proteção de Dados – Comentada**. 2 ed. São Paulo: Thomson Reuters Brasil, 2020. 300 p. ISBN: 9788553219254

MCMAHAN, Brendan; RAMAGE, Daniel. **Federated Learning: Collaborative Machine Learning without Centralized Training Data**. Google: 2017. Disponível em: https://ai.googleblog.com/2017/04/federated-learning-collaborative.html. Acesso em: 15 jan. 2021.

NARAYANAN, Arvind; SHMATIKOV, Vitaly. **Privacy and Security Myths and Fallacies of "Personally Identifiable Information"**. Communications of the ACM. Jun 2020, Vol. 53, No. 08.

_____.; _____. **Robust De-anonymization of Large Sparse Datasets**. Texas: University of Texas at Austin, 2008.

OLIVEIRA, Ricardo de; COTS, Márcio (Coord.). **O Legitimo Interesse e a LGP**D. São Paulo: Thomson Reuters Brasil, 2019 250 p. ISBN: 9786550651770

PATEL, Twinkle; AMIN, Kiran. A Study on k-anonymity, l-diversity, and t-closeness: Techniques of Privacy Preservation Data Publishing. **IJIRST –International Journal for Innovative Research in Science & Technology**, Volume 6, Issue 6, Nov 2019. ISSN (online): 2349-6010. P. 4.

PICCELLI, Roberto Ricomini. **A dimensão política da privacidade no direito brasileiro**. Rio de Janeiro: Lumen Juris, 2018. 171 p., 21 cm. Inclui referências. ISBN 9788551905562.

PILLOW, Timothy. **A Review of Synthetic Tabular Data Tools and Models: Anonymization methods that are revolutionizing how we share data**. Medium: 02 jul. 2020. Disponível em: <https://towardsdatascience.com/a-review-of-synthetic-tabular-data-tools-and-models-d83b232aae25>. Acesso em: 15 nov. 2020.

PINHEIRO, Patricia Peck. **Direito digit**al. 5 ed. rev. atual. ampl. de acordo com as Leis n; 12.735 e 12.737, de 2012. São Paulo: Saraiva, 2013. 438 p. ISBN 978-85-02-20166-8.

_____. **Proteção de Dados Pessoais: comentários à Lei n. 13.709/2018 (LGPD)**. 2ª ed. São Paulo: SaraivaJur, 2020.ISBN 978-85-536-1748-7.

PRIVSEC REPORT. **Data masking: anonymization or pseudonymization?** Disponível em: https://gdpr.report/news/2017/09/28/data-masking-anonymization-pseudonymization>. Acesso em: 08 jan. 2021.

RESENDE, Dirceu. Lei Geral de Proteção de Dados Pessoais (LGPDP ou LGPD) aplicada a bancos de dados SQL Server. Publicado em: 19 mar. 2019. Disponível em: <https://www.dirceuresende.com/blog/lei-geral-de-protecao-de-dados-pessoais-lgpdp-ou-lgpd-aplicada-a-banco-de-dados-sql-server/>. Acesso em: 18 nov. 2020.

RIGOLON KORKMAZ, Maria Regina Detoni Cavalcanti. Dados Sensíveis na Lei Geral de Proteção de Dados Pessoais: mecanismos de tutela para o livre desenvolvimento da personalidade. Dissertação (mestrado acadêmico). Orientador: Sergio Marcos Carvalho de Ávila Negri. Universidade Federal de Juiz de Fora, Faculdade de Direito. Programa de Pós-Graduação em Direito, 2019. 118p.

RODOTÀ, Stefano. A vida na sociedade da vigilância: a privacidade hoje. Rio de Janeiro: Renovar, 2008. 382 p. Inclui notas bibliograficas. ISBN 9788571476882.

SHMATIKOV, Vitaly. k-Anonymity and Other Cluster-Based Methods. University of Texas, CS 380S: Theory and Practice of Secure Systems. Class presentation slides. Disponível em: < https://www.cs.utexas.edu/~shmat/courses/cs380s_fall09/>. Acesso em: 23 jan. 2021.

SINGAPORE (Personal Data Protection Commission). Guide to Basic Data Anonymization Techniques. Publicado em 25 jan. 2018. P. 10. Item 4.a. Disponível em: <https://www.pdpc.gov.sg/help-and-resources/2018/01/guide-to-basic-data-anonymisation-techniques>. Acesso em: 14 jan. 2021

SWEENEY, Latanya. Simple Demographics Often Identify People Uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000. Disponível em: <https://dataprivacylab.org/projects/identifiability/paper1.pdf?fbclid=IwAR2qeD2uXtZmaUUPQzAStNoO1n2pAQpfzzhnDYg4WPN-yhjZ8oTBHmpmhow>. Acesso em: 15 dez. 2020.

TEFFÉ, Chiara Spadaccine de; TEPEDINO, Gustavo. Consentimento e proteção de dados pessoais na LGPD. In: TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro. São Paulo: Ed. RT, 2019. 2 ed. P. 316

TEPEDINO, Gustavo; FRAZÃO, Ana; OLIVA, Milena Donato (Coord.). **Lei Geral de Proteção de Dados Pessoais e suas repercussões no direito brasileiro**. São Paulo: Thomson Reuters Brasil, 2019. 2 ed. 1072 p. ISBN: 9786550654399. P. 118.

UNIÃO EUROPEIA. **Directiva 95/46/CE do Parlamento Europeu e do Conselho, de 24 de Outubro de 1995**. Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=celex%3A31995L0046>. Acesso em: 15 nov. 2020.

_____. **Regulamento (UE) 2016/679 do Parlamento Europeu e do Conselho** (General Data Protection Regulation). Disponível em: <https://eur-lex.europa.eu/legal-content/PT/TXT/HTML/?uri=CELEX:32016R0679>. Acesso em: 15 nov. 2020.

XIAO, Xiaokui; TAO, Yufei. m-Invariance: Towards Privacy Preserving Re-publication of Dynamic Datasets. Hong Kong: Chinese University of Hong Kong, Department of Computer Science and Engineering, 2007.